Human Path Finding in a Semantic Word Game

Frank Marrs[1,*,+], **Mika J. Straka**[2,+], **and Nicole M. Beckage**[3,+]

[1]Colorado State University, Department of Statistics, Fort Collins, CO, USA
[2]IMT School of Advanced Studies Lucca, Piazza San Ponziano 9, 5510 Lucca, Italy
[3]University of Kansas, Department of Electrical Engineering and Computer Science, Lawrence, KS, 66044 USA
[*]frank.marrs@colostate.edu
[+]these authors contributed equally to this work

ABSTRACT

Human decision-making processes are inevitably influenced by mental images, which are triggered by cues such as objects, concepts, and words. The cognitive associations between words can be captured and represented as a semantic network, in which words represent nodes and association between words are edges. How efficiently can humans traverse this word association network to a given target word if, at each moment, they can only make choices about the next word they will visit? Using participant traversal data from the appropriately designed semantic game MindPaths, we show that human players do not rely on guessing, but navigate this association network quite efficiently, finding a specified target word often in the minimal number of steps. We construct models to capture human paths within the MindPaths game. We find that similarity in overall game length is easy to achieve with a trained random walker. We then consider a model of individual choices using Bayesian estimation of transition probabilities conditional on local network connectivity. We find that it is difficult to capture the individual decisions or the decision making process of human players. More complex models driven by semantic similarities and other types of relations between words are necessary to fully understand how individuals choose a particular path within this semantic word game.

Introduction

How well do we know our own language? The ubiquity of the internet and the advent of blogs and social media have led to an exponentially growing quantity of information created by an emergent culture of blogging, liking and commenting. The ability to read and critically question content have become crucial skills for competent internet users. It has long been recognized by artists as well as propagandists likewise that words invoke feelings and cognitive associations which can easily be used and abused. The cognitive connections between words span a semantic association network, a so-called semantic space. The network is malleable by nature and new associations can be formed as regularly applied by, e.g., marketing and election slogans.

In this paper, we analyze the ability of individuals to intentionally represent and navigate a semantic network representation when given a random start and a specific target. By constraining the full complexity of an individual's semantic space to a predefined network, we explore and model whether individual route choices are based on strategic decisions. Previous research on semantic space suggests that people have an intuitive idea about how "far" apart two words are in terms of semantic similarity and that a network is a useful and predictive representation of semantic information[1,2]. Furthermore, experimental evidence has found that the ability to navigate a semantic space is related to cognitive ability more broadly[3,4]. Additionally, vector representation models such as Word2Vec[5] have shown that distance between vectors is strongly related and even predictive of similarity judgments of individuals.

To address the question of semantic space navigation, a type of information foraging, a semantic word game has been conceived in which players have to reach a target word starting from a random word, moving along an underlying semantic network composed of words[6]. This game has further been adapted to an online platform, allowing for more than 800 individuals in many different countries to participate in the game[1]. The underlying network is composed of words as nodes and directed edges, which represent cognitive associations according to the University of South Florida's free association norms database (see Methods). Links can be one-directional, such as *bone* ← *dog* or bidirectional, e.g. *dog* ↔ *cat*, if one word is a response to the cue word and vice versa. Shortest paths from target to end word are not unique and may follow different intuitions, as illustrated by $team \rightarrow football \rightarrow kick \rightarrow scream \rightarrow horror$ and $team \rightarrow coach \rightarrow stage \rightarrow fright \rightarrow horror$.

While it is unlikely that the underlying semantic association network used in this game captures the full richness and uniqueness of an individual's semantic network, performance within this game can nonetheless provide researchers with an understanding about the type of information people access when making a choice of what word to move to next. We

[1]mindpaths.socientize.eu

explore whether players are using some underlying strategy for their decision-making process and whether their choices can be explained in terms of pure network topology or similarity in semantic space. Can such network measures as eigencentrality and in-degree capture how individuals are making specific choices? Or does more informed semantic similarity such as those based on distances in vector space representations better account for performance of individuals in this navigation task? We construct a Bayesian model based on the former assumption and assess our model on the ability to capture and predict individual human choices and collective behavior in this task.

The rest of the paper is organized as follows. We first analyze the overall performance of the human players in the semantic word game and underline how the number of choices to successful navigation increases as the distance between start and end words increases. Subsequently, we introduce a benchmark model for the decision-making modeling, namely an unbiased random walker, and show that the players' strategies are clearly not based on mere guessing. We then investigate a trained random walker model which manages to replicate the general trends of human performance quite well, having a higher success rate on trials that require fewer intermediate word choices. However, we observe a discrepancy regarding individual word choices in particular games. The trained random walker will choose, in the limit, each word that decreases the distance to the target word with equal probability, whereas the human players seem to have a preference for specific paths which take them to the target word. Finally, we introduce a Monte-Carlo Markov-Chain (MCMC) modeling framework which expresses the transition probabilities from one word to another in terms of structural features of the network. For example, our Bayesian model may learn to favor high degree words when the target is far away and then narrow down on the specific goal word by means of eigencentrality. We conclude from the performance of this class of models that different aspects of the network seem to be of more or less importance for the decision-making and for the success of individuals.

Results

Our work is based on the data collected from the word game MindPaths[2]. This game is based on a prior research experiment discussed in Beckage, Butts and Steyvers 2012[7]. In this game, individuals are asked to navigate from a start word to an end word through a series of discrete choices on a predefined semantic network consisting of 2392 unique words and 23,571 directed link. Each link represents a free association between two words (see Methods). The distance from start to end word could take a value between one, in which the player can directly select the target word, to a maximum of six. Only trials where individuals reach the end word in less than 25 steps are included in our analysis, since reasons for unsuccessful plays are many faceted and can include uncontrollable factors such interruption of internet connection. See[7] for an analysis exploring the overall rate of success in this type of semantic game. For interested readers, other work has focused explicitly on why individuals give up in a similar network navigation task[8]. The original data contains information on the decision times in milliseconds, which we neglect here but which offers possibilities for future research.

Game design and Vocabulary

For ease of reference to important states of the game, we briefly introduce a few terms. The *target word* refers to the word that an individual is trying to find. If this word is selected the game ends. The *geodesic distance* between two nodes is the minimum number of steps required to reach one node starting from the other, i.e. the length of a shortest path between these nodes on the network. The total number of moves which a player uses to solve a game is referred to as the *game length*. We define the *current word* as either the start word or the most recent word selected by an individual. The *option set* are the neighbors that the current word points to. The option set is constrained to be between 3 and 12 words. If a particular word had more than 12 out-neighbors, the strongest 12 were presented. Words with less than 3 associates were not included in the game network. We define the *choice* to be the specific word from the option set that was selected by an individual or a model. Fig. 1 shows a screen shot of the game.

Human Performance

As shown in Fig. 2, the players perform surprisingly well in the semantic games, solving games in the minimum number of steps in approximately 33% of the successful trials. This indicates that they were actively searching to reach the target word quickly and that words lying on the shortest paths met their ideas of being effective moves. In some cases, however, players require significantly more steps. The increase in game length for particular individuals or games could be due to a simple guessing strategy or to the fact that the network representation is not in line with the players' expectations. In fact, navigation to some of the targets along the geodesic do sometime require topic shifts. One such example is the example from above where individuals search for a path of *team* → ... → *horror*. In this case, the start and end words are thematically very different. Another common feature of these games is that the shortest path may require the use of polysemy, or the use of two different meanings of the same orthographic word, to alter the concept space as in the play *pillow* → *sheet* → *paper* → *pen*. A general

[2]See mindpaths.socientize.eu for more detail, to play the game and to contribute to ongoing research.

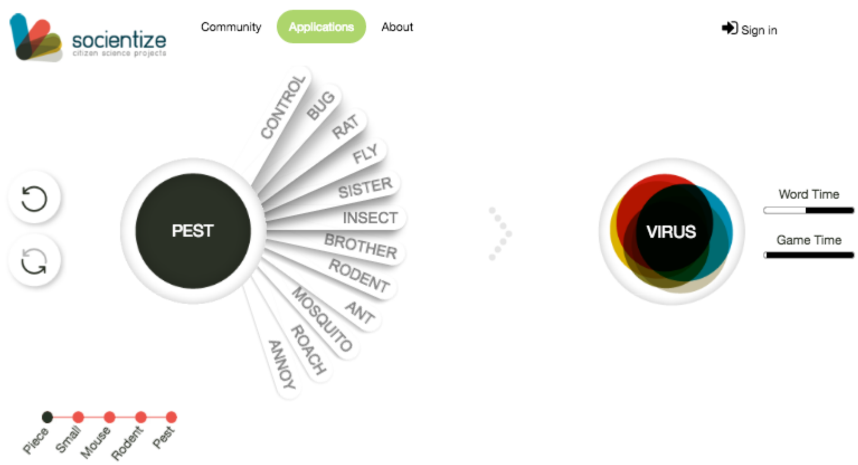


Figure 1. Screen shot of the online game. The word in the left circle is the current word, the word in the right circle is the target word. Individuals must select one word of the option set (protrusions of the left circle) with the aim of moving 'closer' to the target word. The selected word (or choice) becomes the current word in the next update.

trend is that the players' ability to reach the goal along the geodesic decays significantly as the game geodesic between start and target word increases, see Fig. 2.

Random Walker Models

While navigating through the semantic network, at each stage the player only knows the current word she is at, the option set for the next move, and the target word. She also has access to the start word and the sequence of all words previously selected in reaching the current state, but this information is not necessarily relevant since the purpose of the game is to move closer to the target word. Furthermore, each transition occurs with some amount of uncertainty because the player does not know whether the new word will really bring her closer to the target due to her lack of knowledge of the full underlying network structure. The player therefore may be approximated as a memoryless Markov process. Thus for our baseline model we compare human performance to an unbiased random walker (RW) as a worst-case scenario benchmark. This model will mimic human performance if individuals are simply guessing or choosing an option randomly. We also consider a biased random walker with dynamic transition probabilities which are iteratively updated based on paths that successfully reach the target word. From this, we subsequently build a more cognitively-based generative model that aims to account for the decision process underlying human choices in this semantic network game.

Unbiased Random Walker

Our previous analysis of players' successful games has shown that humans navigate rather quickly and effectively within the network, especially for games of short geodesic distances. An unbiased random walker serves as a benchmark to test whether a pure guessing strategy of the players can be excluded. In fact, the random walker chooses uniformly from the options set of the current word unless the target word is included. Note that the certain selection of the target word is notably better than the humans, as humans will choose another option even when the target word is present in about 6% of cases. For the analysis, only the path lengths of successful random walk trials are taken into consideration.

To generate the random walker performance, we initialize the walker with the same specific games as those played by humans. We consider a trial a success if, through uniform selection, the random walker reaches the target word in 25 steps or less. We repeat the process for each trial, proportional to the number of games actually played, for totals between 100 and 5,400 repetitions of each game. The results are summarized in Fig. 3. It shows that humans outperform unbiased RW at all game lengths, confirming that the human decision strategies are more sophisticated than random guessing. We conclude that human players are using their understanding of semantic space to aid in navigation to the target word.

Trained Random Walker

In order to simulate the performance of human players, we train a random walker (TRW) by generating a distribution for word transition probabilities conditional on the target word based only on game geodesics. Explicitly, we seek $\phi_{jk}|t$, the probability of selecting word k given the player is at word j and is seeking target word t. For the complete algorithm, see Alg. 1 in the

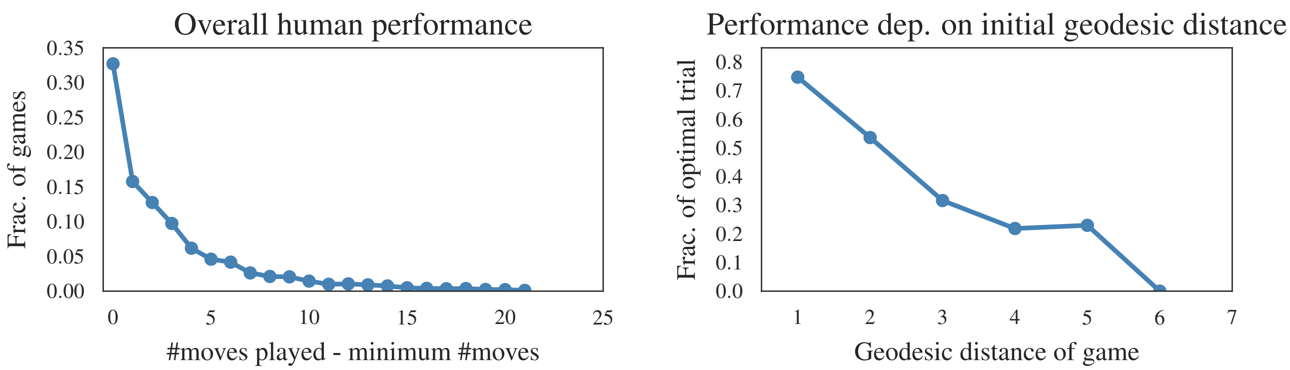


Figure 2. **Left:** Overall performance of the players. The figure show how often players have solved the games in the minimum number of steps necessary and how often they have needed more, independent from the initial geodesic distance of the games. In about 33% of all cases, the games have been solved in the optimal number of steps. **Right:** Fraction of optimally solved games depending on the geodesic distance between the start and the target word. Whereas over 50% of all games of lengths two or one are solved optimally, the ratio decays quickly for longer games.

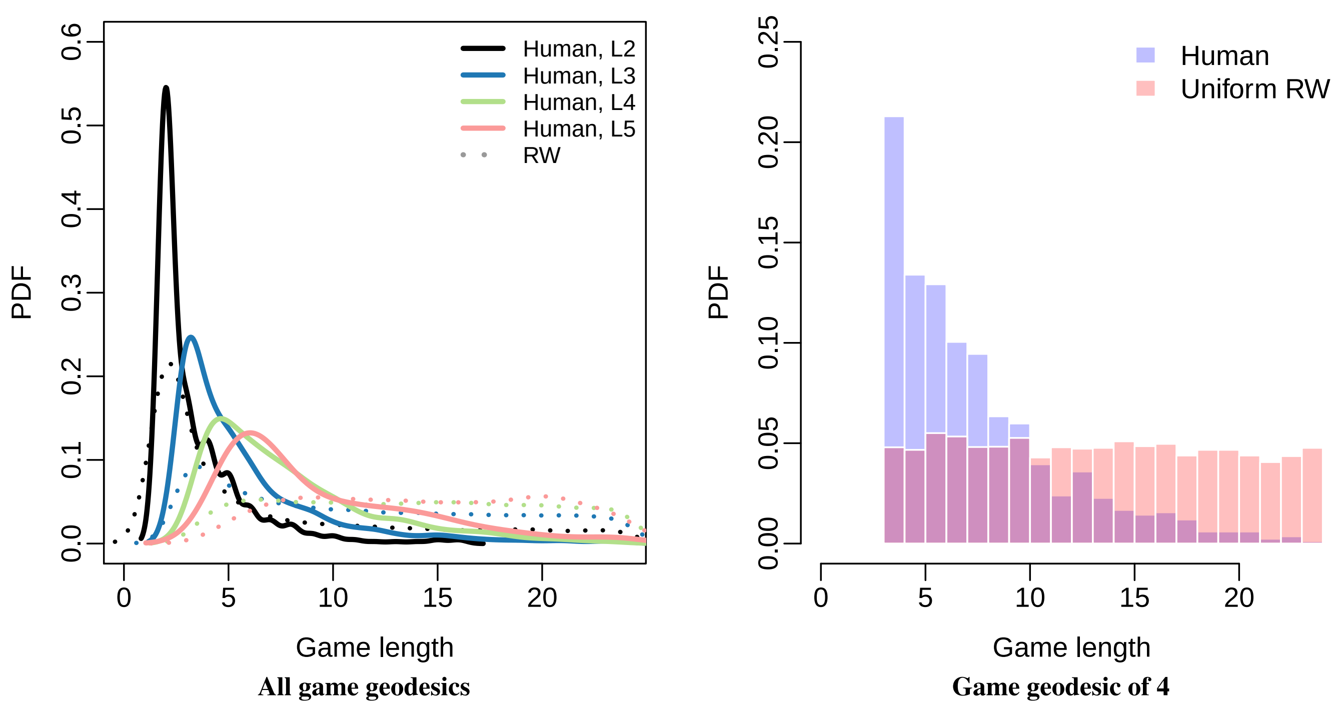
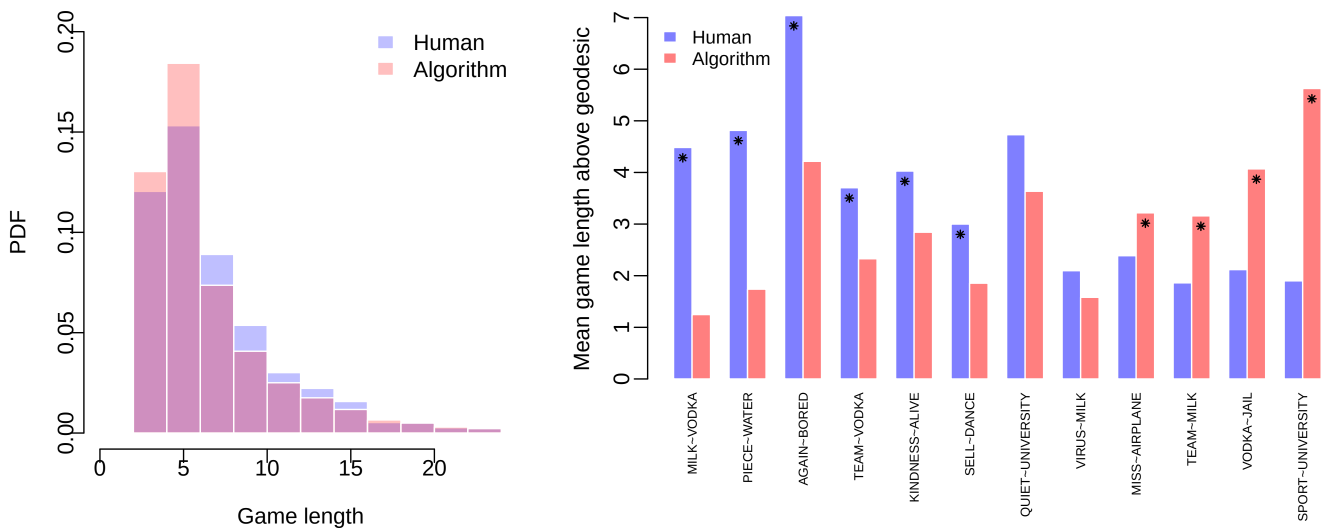


Figure 3. Performance of players compared to a random walk for games of geodesic length 2 through 5, and game geodesic 4 in particular. The discrete probability density functions in the left panel are smoothed using Gaussian kernel. In all cases, human players clearly outperform the random walker.

Methods section. To emphasize, the purpose of training the random walker is to see if we can emulate the human learning process with only experience in the games and without further information regarding the underlying network.

Before each update of the transition probabilities, we executed 500 random walks. Since the performance appears to converge as the number of updates grows, we settled for 80 updates resulting in a total of 40,000 runs. Because convergence of transitional probabilities is target word specific, we concentrated on the ten most commonly played games of geodesic length 4 as length 4 is the most common game. We assessed the TRW's performance by starting 2,000 simulated games at each start word and taking random steps according to $\phi_{kl}|t$ until the target word was reached or until 25 steps were taken. As in the human performance data, we discarded walks that took more 25 steps (less than 5% of all walks after initial training). The results are summarized in Fig. 4.

The performance of the algorithm in the aggregate matches that of the humans quite well, as shown in the left panel. However, there is significant variability between human and algorithm performance when examining each game individually (right panel of Fig. 4). If we assume that the game length above geodesic is distributed Poisson(λ), then the sample mean game length above geodesic is the maximum likelihood estimator of λ. Further, most of these sample means are significantly different (indicated by "$*$") at the .05 level using the conditional test of Poisson parameters[9]. We find similar results when employing a nonparametric test[10]. The variability among games in the random walker performance suggests that human players are influenced by information beyond just the shortest possible path from start word to target word. Further, experience with the network representation directly is not enough to capture human performance in this task.



All games together **Separate games**

Figure 4. Performance of players compared to trained random walker (Algorithm). Games are of geodesic 4, using only those games that end in the 10 most common target words. "$*$" indicates significantly different mean values at level .05.

Modeling semantic transition probabilities from network observables

The transition probabilities of passing from one node to the next in the semantic network should ideally reflect the cognitive decision-making process of the players. In the random walker analysis we showed that similar decisions can be made by assuming a trained random walker which is updated based on successful transitions. However, we also see that this model is only able to account for the aggregate performance of human players and fails to account for the variably on a specific trial. The trained random walker is the result of repeated training repetitions over the association network, and hence is not generative. It can be argued that the trained random walker simply "learns" the association network. We seek to develop a generative identify of player performance in this section that can help to model the types of information used by individuals when performing semantic network navigation.

We ask whether it is cognitively plausible that individuals are choosing to transition from the current word to a specific option word based on the network connectivity of the words. We model the player choices as defining the transition probabilities conditioned on the connectivity of a specific word. For example, individuals might begin their initial search by trying to find a word that is similar in theme to the target word. One explicit strategy might be to choose words that are likely to have many associates in the hopes of finding a thematic word in the option set. Within this model, individuals are more likely to select high degree words the farther they are from the target word. Other types of network measures or vector distance metrics might be equally relevant to an individuals strategy in searching for the target word. We initially focus on network connectivity to give insight into the navigation strategies of players within the network. Let $\mathbf{Q}_{kl}$ be a vector of such network quantity measured on the nodes couple $\{k,l\}$, for example $\mathbf{Q}^i_{kl}$ may indicate the in-degree of node l and $\mathbf{Q}^{i+1}_{k,l}$ may indicate the distance between node k and l in the underlying graph. $\theta \in \mathbb{R}^p$ is a vector of weight parameters, capturing the influence of each $\mathbf{Q}_{kl}$ element on the decision process of an individual. In general, the transition probabilities will be some function on these quantities,

$$\phi_{kl} = f\left(\{\mathbf{Q}_{kl}, \theta\}\right). \tag{1}$$

Transitions from node k to node l are only possible if there is a link from k to l and probabilities must sum to one. A straightforward probability model is given by

$$\phi_{kl} = \frac{exp(\theta^T \mathbf{Q}_{kl})}{\sum_i exp(\theta^T \mathbf{Q}_{ki})}, \tag{2}$$

The sum in the denominator runs over all neighbors i reachable from k. In first approximation we assume that the probability model is not conditional on the target word. Notice that $\theta = \mathbf{0}$ recovers the unbiased random walker discussed previously. Since the expression (2) is invariant to the addition of a constant to the argument, i.e. $f(\{\mathbf{Q}_{kl}, \theta\}) = f(\{\mathbf{Q}_{kl}, \theta + c\})$ for any $c \in \mathbb{R}^p$, no bias/intercept term is required.

In general, many different network measures for $\mathbf{Q}_{kl}$, such as in-degree, out-degree, betweenness and eigencentrality, could be of interest for our probability model. To help with its evaluation, we introduce the idea of good, bad, or neutral choices of a particular player, which either decrease, increase, or neither increase nor decrease her distance to the target word. We note that usually every move type is performed under the assumption of taking the player closer to the goal. This allows us to understand which network features are both salient and meaningful to human players in successfully (or even optimally) completing a particular game. By calculating correlations of parameter estimations with good or bad player choices in real games, we found eigencentrality and in-degree most relevant and interpretable for our analysis (see Methods for details).

The distribution of θ was estimated via Markov-Chain Monte-Carlo (MCMC) with a mean-zero prior for θ and updated using a Metropolis algorithm[11] (see Methods). We consider three different network measures $\mathbf{Q}$, namely examine eigencentrality alone, eigencentrality and in-degree together, and eigencentrality with a different weight parameter for each distance from the target word. In all estimation procedures, we choose the uninformative prior standard deviation of $\tau = 100$. The proposal standard deviation was tuned such that the acceptance rate of the Metropolis algorithm was between approximately 0.2 and 0.5[12]. The estimation chains were run until there were at least 1,000 effective samples for each parameter of θ. For more details, see additional information in the Appendix.

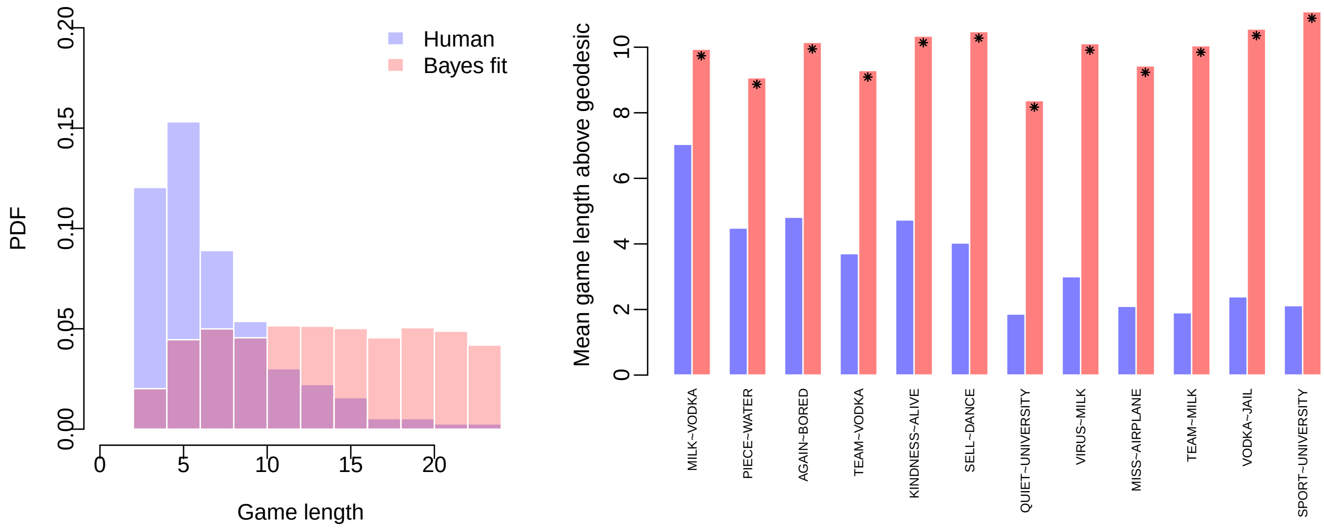
We assess model performance with the Akaike Information Criterion (AIC)[13], using the likelihood $p(\hat{\theta}|\mathbf{y}, \mathbf{Q})$ of observing parameters $\hat{\theta}$ given the players' moves $\mathbf{y}$ and the network measures $\mathbf{Q}$. AIC evaluates the model goodness of fit while penalizing for complexity, i.e. the number of fitted parameters p. We also calculate the Kullback-Leibler (KL) divergence from the estimated probability distribution to the observed distribution[14]. Note that the observed distribution is an estimate of the true underlying distribution, and the KL divergence calculated is really the sum of many KL divergences, since we have a unique estimate for each word in the network. Further, note that AIC is an asymptotically consistent estimator of the KL divergence between the estimated distribution and the true underlying distribution.

Table 1. Measures of model fit resulting from MCMC estimation of θ. p is the number of parameters and KL is the Kullback-Leibler divergence.

Components of $\mathbf{Q}$	p	AIC	KL in-sample	KL out-of-sample
Completely random walker	0	15119.7	.455	.734
Eigencentrality only	1	14727.3	.438	.720
Eigencentrality and In-degree	2	14722.4	.438	.719
Eigencentrality by geodesic	4	14698.2	.437	.719

In examining Table 1, we find that each of the fitted models outperforms the unbiased random walker baseline. Based on KL divergence, there is no difference among the fitted models. However, a difference of 2 units in AIC is significant, which implies that each model is an improvement over the previous. As quantified by AIC, the best model uses the most parameters; the model that fits eigencentrality conditioned on geodesic. This fact suggests that incorporating further measures in $\mathbf{Q}_{kl}$ and specifying separate parameters for each geodesic may improve fit of the Bayesian model further.

We repeat the analysis of Figure 4 for the trained random walker on the best-fitting Bayesian model (that based on a different eigen-centrality parameter for each geodesic). We assess the Bayesian model's performance by starting 2,000 simulated games at each start word and taking random steps according to the estimated probability of ϕ_{kl} from (2) until the target word was reached or until 25 steps were taken. As in the human performance data, we discarded walks that took 25 steps (as much as 95% of all walks). The results are plotted in Figure 5. We find that the performance of the Bayesian model looks quite similar to that of the uniform random walker in Figure 3 and not nearly as good as that of the trained random walker in Figure 4. This result is not altogether surprising, as the trained walker is based on many training repetitions over the network while the Bayesian model is only based on highly localized network measures. While the Bayesian model clearly provides some predictive power over uniform probabilities as evidenced by Table 1, there remains significant variation in the players choices for which to be accounted.



All games together **Separate games**

Figure 5. Performance of human players compared to the best-fitting Bayesian model (labeled "Bayes fit"). Games are of geodesic 4, using only those games that end in the 10 most common target words. "∗" indicates significantly different mean values at level .05.

Discussion

In the present article we have addressed the question of how efficiently humans can navigate a semantic space in the form of a game on a word association network. Our analysis has shown that players do not rely on simple guessing but are relatively good at searching for the shortest possible paths from the start to the target word. We also find that their choices are not only influenced by distance minimization, but also by possible network measures or semantic relations between words. We modeled the decision making process of individuals by considering individual word transition probabilities to be functions of the eigencentrality and in-degree of the current option sets. Both quantities are plausible as cognitively important and locally available features. These measures can be thought of as a systematic way to express the preference of choosing nodes which give the player an higher probabilities to continue advancing in the direction of the target node. Our results show that an eigencentrality model, in which the probability parameters are weighted differently depending on the distance from the target, performs best among the tested models, suggesting that strategies of individuals may change based on perceived closeness to the target word. However, in general we have shown that a simple model based on one or two network quantities does not sufficiently explain the cognitive decision-making processes of the players. Further investigation of the contribution of other word specific measures and similarity or distance between a current word and the target word may improve our models. In the future we hope to better account for individual choices as well as to offer cognitive and decision explanations to capture what features of semantic space are available to individuals in this task.

The Bayesian framework introduced here is easily extendable to include other word features. One obvious extension would be to consider features of the words such as the word frequency and part of speech. Additionally, we are interested to include semantic similarities between words as well, especially similarity between the current and target word. In fact, latent spaces such as Word2Vec[5] offer another layer of information which promises to shed light on the decision-making process of the players. Further, network measures that are conditional on the target word, as is the trained random walker, may provide significant improvements in the performance of the generative model.

In future work, we plan to expand the possible local features that may influence a particular choice beyond our two measures of network centrality. However, since network measures are highly correlated, we must first construct informed priors or employ a regularization technique that allow us to uncover the underlying decision process of individuals directly. Extending our Bayesian models with other semantic features that may influence an individual's choice of words to move to next may help us understand how individuals integrate local and global information when foraging for information. We saw here that only local information is not sufficient to account for individual choices. Future work will focus on what types of global information is available, how is it stored and most importantly how are people using it to solve this network navigation task?

Methods

The Word Associations Network

Associations between words may differ from individual to individual, but it is reasonable to assume that some underlying network structure exists, at least within certain cultural and linguistic borders. In order to sample the "average" network, Nelson et al. collected the University of South Florida Free Association Norms comprising a database of 5,019 normed words and data on 72,000 word pairs[15]. In this study individuals are asked to respond with the first semantically similar word that comes to mind when hearing a cue word. For example, upon hearing the word *dog*, 64% of the participant in the survey have answered with *cat*. The frequency of response can be considered as an edge weight in a free-association network.

Our semantic game uses a subset of this original free-association network. In particular, words need to have at least an out-degree of three to be included and can have a maximum out-degree of twelve. As a consequence, our directed network reduces to 2,392 unique words and 23,571 binary and unweighted edges. Note that associations do not have to be symmetric, meaning that edges are, in general, one-direction. The connectance of the directed network is about 0.4% and the diameter 8.

The Semantic Game

In the game, the player starts in an arbitrary location in the word association network described above and must navigate, via forced choice, to a target word. At each moment, there will be at least one shortest path from the current player's position to the target word. When the player moves from one word to another, the shortest distance to the target word either increases, remains constant, or decreases based on what can be considered a negative, neutral or positive choice. A trial can comprise not more than 25 moves in order to be considered successful. Geodesic distances between start and end words reach from one, where the target word is already included in the option set of the start word, to six. The layout of the game is illustrated in Fig. 1.

Trained random walker algorithm

The transition probabilities of the trained random walker are updated according to the algorithm illustrated in Alg.1. After a particular number of games U were completed, the transition probabilities were updated. If a particular path resulted in correct selection of the target word, transition probabilities along that path were augmented proportional to the inverse path length. Transition probabilities were then renormalized. This process was repeated a total of W times. We tested a range of different numbers of updates $U \in \{5, 10, 20, 40, 80\}$ and observed stability of the RW's performance with $U = 80$ updates. All analysis is based on this case. In total, 40,000 random walks were executed.

for $u \in \{1,2,\ldots,U\}$ **do**
  for $w \in \{1,2,\ldots,W\}$ **do**
    i. select start word uniformly at random.
    ii. take steps completely at random until the target word t is reached or 25 steps are taken.
  end
  update $\phi_{jk}|t \propto \sum_u \sum_w \frac{1}{L_{jk}^{(u,w)}}$ and normalize appropriately.
end

Algorithm 1: Algorithm for training the random walker with probabilities based on inverse weighting of game length. $L_{jk}^{(u,w)}$ is the length of test walk w in update step u. We define $L_{jk}^{(u,w)} := \infty$ when the edge (j,k) is not traversed in test walk (u,w) or when the length of the test walk reaches 25.

Bayesian MCMC

We estimate distribution of θ via Markov-Chain Monte-Carlo (MCMC). More specifically, we seek an estimate of the posterior distribution

$$p(\theta|\mathbf{y},\mathbf{Q}) = g(\mathbf{y}|\theta,\mathbf{Q})h(\theta), \tag{3}$$

where $g(.)$ is the likelihood of the data given the parameters, $h(.)$ is a prior distribution on θ, and $\mathbf{y}$ and $\mathbf{Q}$ are the observed player choices and the set of network parameters, respectively. A natural choice of likelihood is a multinomial distribution, given that a player is at word k, she must choose a single step from a set of options. Let there be L_k options for word k and K total words. Then we may write the likelihood

$$g(\mathbf{y}|\theta,\mathbf{Q}) = \prod_{k=1}^{K}\left(\sum_{l=1}^{L_k} y_{kl}\right)!\prod_{l=1}^{L_k}\frac{\phi_{kl}^{y_{kl}}}{y_{kl}!}, \tag{4}$$

where y_{kl} is the number of times edge (k,l) is traversed in the dataset and ϕ_{kl} is as given in (2). Note that (4) makes the implicit assumption that each step is independent of every other.

The Dirichlet distribution for ϕ_{kl} is conjugate the likelihood in (4). However, since we propose a model for ϕ_{kl} (rather than inferring a separate value for each combination of k and l), there is no conjugate distribution. Thus, we choose a mean-zero prior for θ such that $h(\theta) \sim N(0, \tau^2 I_p)$.

In the absence of conjugacy, we must use the Metropolis algorithm[11]. The estimation procedure is as follows:

1. Given previous sample $\theta^{(s)}$, propose $\theta^* = \theta^{(s)} + \delta$, for random $\delta \sim N(0, \varepsilon^2 I_p)$
2. Evaluate $p(\theta^*|\mathbf{y}, \mathbf{Q})$
3. Accept the proposal (i.e. let $\theta^{(s+1)} = \theta^*$) with probability $\frac{p(\theta^*|\mathbf{y},\mathbf{Q})}{p(\theta^{(s)}|\mathbf{y},\mathbf{Q})}$. Otherwise, generate a new proposal and set $\theta^{(s+1)} = \theta^{(s)}$.

Network Measures for Bayesian Estimation

We consider only a subset of possible measures for $\mathbf{Q}$ as defined in (3). These measures are based on data exploration comparing the choices made by participants and the choices correlation with normalized network measures. Let us consider a game where the player resides at node k and selects node l. For example, we compute $\frac{exp(d_{kl})}{exp(\sum_i d_{ki})}$, where d_{kl} is the out-degree of node l that is connected to k. We compare this computation to the observed proportion of selections of node l when players reside at node k via Kendall rank correlation. In general, we find a small but significant correlation between all examined network measures and the observed proportions ($p < 0.1$ for in-degree and $p < 1 \times 10^{-7}$ for all other measures). The correlations are summarized in Table 2.

Table 2. Kendall rank correlation between network measures and observed proportions of selections. "centr." is an abbreviation for centrality.

In-degree	Out-degree	Betweenness	Closeness centr.	Harmonic centr.	Eigencentr.
0.010	0.045	0.036	0.042	0.049	0.046

We also examine the correlation within the network measures reported in Table 2. Unsurprisingly, we find large correlations between the network measures for this data set. In fact, many of the correlations are above .75, indicating that there is little additional information in each network measure when one network measure is already known. We choose eigencentrality and in-degree for further investigation. The choice of the former is due to its large correlation value; in-degree has the minimum absolute correlation with eigencentrality. All the remaining measures have correlation greater than 0.75 with at least one of eigencentrality or in-degree.

References

1. Collins, A. & Quillian, M. Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Verbal Behavior* **247**, 240–247 (1969).

2. Collins, A. & Loftus, E. A Spreading-Activation Theory of Semantic Processing. *Psychological Review* **82**, 407–428 (1975).

3. Goñi, J. *et al.* The semantic organization of the animal category: evidence from semantic verbal fluency and network theory. *Cognitive Processing* **12**, 183–96 (2011).

4. Borge-Holthoefer, J., Moreno, Y. & Arenas, A. *PloS ONE* .

5. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. *CoRR* **abs/1301.3781** (2013). URL http://arxiv.org/abs/1301.3781.

6. Beckage, N., Steyvers, M. & Butts, C. Route choice in individuals—semantic network navigation. In Miyake, N., Peebles, D. & Cooper, R. (eds.) *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 108–113 (Cognitive Science Society, Austin, T.X., 2012).

7. Beckage, N. M., Steyvers, M. & Butts, C. T. Route choice in individuals—semantic network navigation. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 108–113 (2011).

8. Scaria, A. T., Philip, R. M., West, R. & Leskovec, J. The last click: Why users give up information network navigation. In *Proceedings of the 7th ACM international conference on Web search and data mining*, 213–222 (ACM, 2014).

9. Przyborowski, J. & Wilenski, H. Homogeneity of results in testing samples from poisson series: With an application to testing clover seed for dodder. *Biometrika* **31**, 313–323 (1940).

10. Arnold, T. B. & Emerson, J. W. Nonparametric goodness-of-fit tests for discrete null distributions. *The R Journal* **3**, 34–39 (2011).

11. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**, 1087–1092 (1953).

12. Hoff, P. D. *A first course in Bayesian statistical methods* (Springer Science & Business Media, 2009).

13. Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723 (1974).

14. Kullback, S. & Leibler, R. A. On information and sufficiency. *The Annals of Mathematical Statistics* **22**, 79–86 (1951).

15. Nelson, D., McEvoy, C. & Schrieber, T. The university of south florida word association norms (1999).

Acknowledgements

We thank the participants of the 2016 Santa Fe Institute Complex Systems Summer School for their feedback and insightful discussion.

Author contributions statement

N. B. designed the original game and provided access to the data, F. M. conducted most of the data analysis and modeling, M. S. drafted the manuscript, everybody discussed at El Farol.

Appendix: MCMC Estimation Details

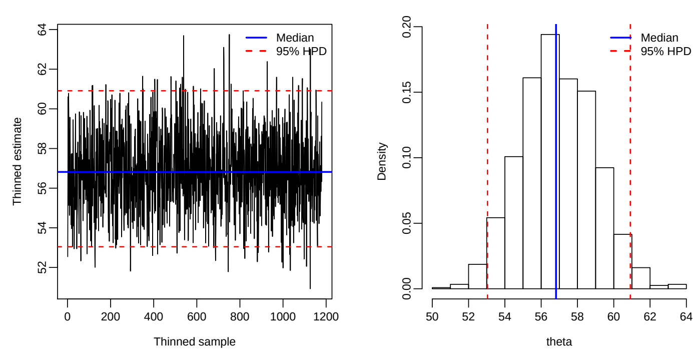
Here we detail some of the choices made to complete the MCMC estimation. As previously noted, a diffuse prior with standard deviation chosen was $\tau = 100$ for all models and parameters. In what follows, we address each model individually.

Univariate model

We selected eigencentrality as the most significant network measure. The proposal standard deviation chosen was $\varepsilon = 9$. We find good mixing when using this network measure and ε: 2,360 proposals led to 1,180 effective samples after thinning at a rate of every other accepted proposal. The coefficient estimates are given in Table 3. The thinned MCMC chain and posterior distribution estimate are given in Figure 6. We see that the coefficient estimate for θ is significantly nonzero based on the 95% highest probability density (HPD) bounds.

Table 3. Univariate MCMC results (eigencentrality only network measure).

Coefficient θ	Median	95% lower bound	95% upper bound
eigencentrality	56.8	53.0	60.9



MCMC chain **Posterior distribution**

Figure 6. Univariate MCMC results (eigencentrality only network measure).

Univariate model with distance-specific parameters

We again used eigencentrality as the most significant network measure, but estimate a different parameter for each geodesic from the target word $\{2,3,4,5\}$, such that $\theta \in \mathbb{R}^4$. The proposal standard deviation chosen was $\varepsilon = 5$. We find moderate mixing when using this network measure and ε: 42,954 proposals led to a minimum of 1,045 effective samples after thinning every 35. The coefficient estimates are given in Table 4. The thinned MCMC chain and posterior distribution estimate are given in Figure 7. All the coefficients are significantly different from zero, and all are significantly different from each other except for those pertaining to geodesics 4 and 5.

Table 4. Univariate MCMC results for eigencentrality with separate geodesic-specific coefficients.

Coefficient θ	Median	95% lower bound	95% upper bound
eigencentrality, geodesic 2	78.6	71.4	85.7
eigencentrality, geodesic 3	55.2	49.0	61.8
eigencentrality, geodesic 4	29.3	21.5	37.1
eigencentrality, geodesic 5	30.2	15.6	43.4

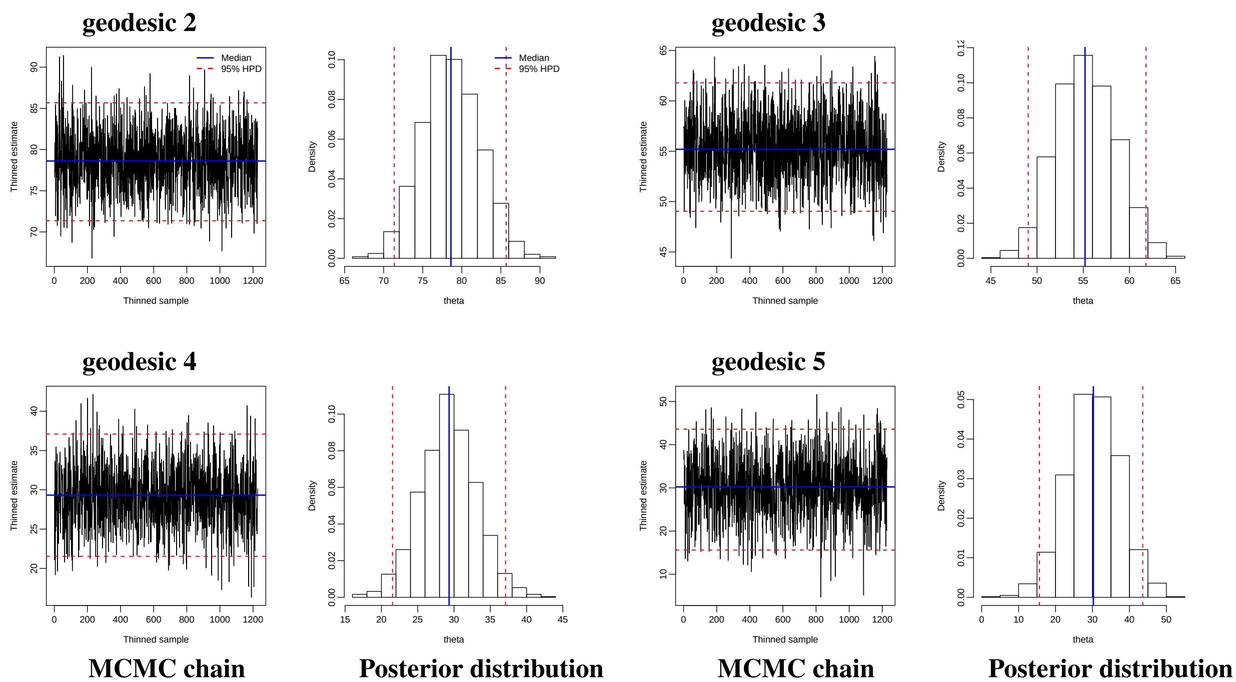


Figure 7. Univariate MCMC results for eigencentrality with separate geodesic-specific coefficients.

Bivariate model

We used eigencentrality and in-degree as network measures in $\mathbf{Q}$, such that $\theta \in \mathbb{R}^2$. To assess the relative importance of these measures and to ensure effective model fit, we standardized these network measures such that they each had standard deviation 1. The proposal standard deviation chosen was $\varepsilon = .015$. We find relatively poor mixing when using this network measure and ε: 137,351 proposals led to at least 1,281 effective samples after thinning every 40. The coefficient estimates are given in Table 5. The thinned MCMC chains and posterior distribution estimate are given in Figure 8. Table 5 and Figure 8 present the rescaled values of θ that may be directly multiplied by unscaled network measures in-degree and eigencentrality, respectively. Again, we observe that both θ coefficients are significantly nonzero based on the 95% highest probability density (HPD) bounds. Further, in the scaled model, the eigencentrality HPD interval $[0.28, 0.32]$ is significantly greater than the in-degree HPD interval $[0.026, 0.060]$. This fact suggests that eigencentrality is more predictive of player choices than in-degree.

Table 5. Bivariate MCMC results. Presented coefficients are rescaled.

Coefficient θ	Median	95% lower bound	95% upper bound
eigencentrality	53.6	49.9	57.6
in-degree	0.00155	0.000936	0.00219

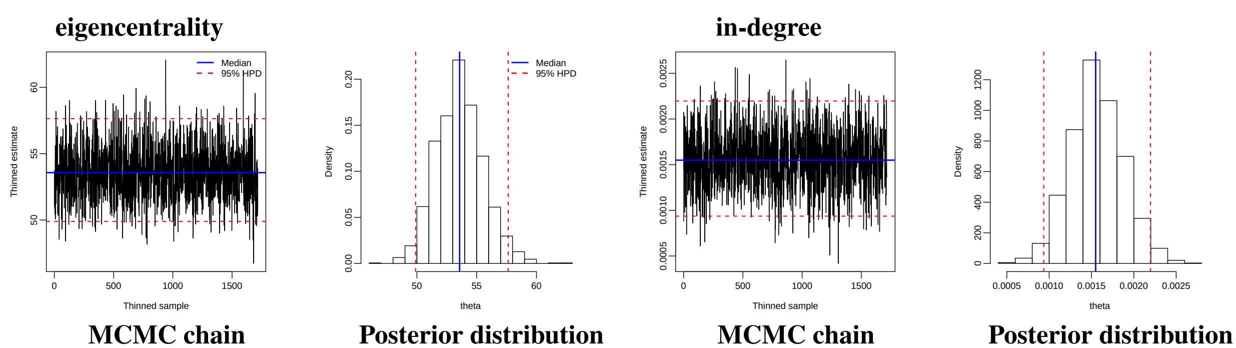


Figure 8. Bivariate MCMC results. Presented coefficients are rescaled.